DOCUMENT RESUME

ED 443 834 TM 031 315

AUTHOR Onwuegbuzie, Anthony J.
TITLE Integration of the Rival Hypotheses Tool into Research Methodology Courses: Issues and Strategies To Support Its Use and Sustainability.
PUB DATE 2000-04-25
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Opinion Papers (120) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comprehension; *Hypothesis Testing; *Research Methodology; *Validity; *World Wide Web

ABSTRACT

This paper provides a rationale for using and sustaining rival hypotheses Web-based tools to promote students' understanding of the concepts of internal and external validity. Five major concerns are identified. The first is that, in their present form, the Web sites subsume the discussion of threats to validity under experimental designs, giving the impression to some students that such threats are not an issue for other types of quantitative research. The second concern is the fact that the illustrative vignettes are presented in multiple-choice formats, giving the impression that each research study has only one threat to internal or external validity, which is an unrealistic assumption. In receiving immediate feedback (i.e., solutions), some students may not reflect deeply enough about the scenarios, preferring to select a response hastily to obtain early validation. In such cases, the critical thinking process involved in the rival hypothesis reasoning will be stunted. Fourth, although analyzing vignettes is an extremely useful exercise, it should be remembered that these vignettes represent mere isolated fragments of information, typically devoid of any theoretical framework. Finally, providing only Web-based tools for teaching the concept of validity with respect to empirical studies may give graduate students and researchers alike the false impression that validity is not an issue in qualitative designs. Recommendations are provided in light of these concerns. (Contains 25 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 443 834

Running head: INTEGRATION OF THE RIVAL HYPOTHESES TOOL

Integration of the Rival Hypotheses Tool Into Research Methodology Courses: Issues and Strategies to Support Its Use and Sustainability

Anthony J. Onwuegbuzie

Valdosta State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A. Onwuegbuzie

1

TM031316

Symposium conducted at the annual conference of the American Educational Research Association, New Orleans, LA, April 25, 2000.

BEST COPY AVAILABLE

Abstract

The purpose of this presentation is to provide a rationale for using and sustaining rival hypotheses web-based tools to promote students' understanding of the concepts of internal and external validity. In so doing, five major concerns are identified. First, in their present form, the websites subsume the discussion of threats to validity under experimental designs, thereby giving the impression to some students that such threats are not an issue for other types of quantitative research. Second, the fact that the illustrative vignettes are presented in multiple-choice formats also gives the impression that each research study has only one threat to internal or external validity, which is an unrealistic assumption. Third, in receiving immediate feedback (i.e., solutions), some students may not be reflect deeply enough about the scenarios, preferring to select a response hastily in order to obtain early validation. In such cases, the critical thinking process involved in the rival hypothesis reasoning will be stunted. Fourth, although analyzing vignettes is an extremely useful exercise, it should be remembered that these vignettes represent mere isolated fragments of information, typically devoid of any theoretical framework. Finally, providing only web-based tools for teaching the concept of validity with respect to empirical studies may give graduate students and researchers alike the false impression that validity is not an issue in qualitative designs. Recommendations are provided in light of these concerns.

Integration of the Rival Hypotheses Tool Into Research Methodology Courses: Issues and Strategies to Support Its Use and Sustainability

Recently, the Committee on Professional Ethics of the American Statistical Association (ASA) addressed the following eight general topic areas relating to ethical guidelines for statistical practice: (1) professionalism; (2) responsibilities for funders, clients, and employers; (3) responsibilities in publications and testimony; (4) responsibilities to research subjects; (5) responsibilities to research team colleagues; (6) responsibilities to other statisticians or statistical practitioners; (7) responsibilities regarding allegations of misconduct; and (8) responsibilities of employers, including organizations, individuals, attorneys, or other clients utilizing statistical practitioners. With respect to *responsibilities in publications and testimony*, the Committee stated the following:

> (6) Account for all data considered in a study and explain sample(s) actually used.
>
> (7) Report the sources and assessed adequacy of the data.
>
> (8) Clearly and fully report the steps taken to guard validity.
>
> (9) Where appropriate, address potential confounding variables not included in the study. (The American Statistical Association, 1999, p. 4)

Although the ASA Committee on Professional Ethics did not directly refer to these concepts, it would appear that these recommendations are related to internal and external validity.

At the same time, the ASA Committee was presenting its guidelines, the American Psychological Association (APA) Board of Scientific Affairs, who convened a committee

called the Task Force on Statistical Inference, was providing recommendations for the use of statistical methods (Wilkinson and the Task Force on Statistical Inference, 1999). Useful recommendations were furnished by the Task Force in the areas of design, population, sample, assignment (i.e., random assignment and nonrandom assignment), measurement (i.e., variables, instruments, procedure, and power and sample size), results (complications), analysis (i.e., choosing a minimally sufficient analysis, computer programs, assumptions, hypothesis tests, effect sizes, interval estimates, multiplicities, causality, tables and figures), and discussion (i.e., interpretation and conclusions).

Although the APA Task Force stated that "This report is concerned with the use of statistical methods only and is not meant as an assessment of research methods in general" (Wilkinson and the Task Force on Statistical Inference, 1999, p. 2), it is somewhat surprising that internal and external validity was mentioned directly only once. Specifically, when discussing the reporting of instruments, the task force declared:

> There are many methods for constructing instruments and psychometrically validating scores from such measures. Traditional true-score theory and item-response test theory provide appropriate frameworks for assessing reliability and *internal validity*. Signal detection theory and various coefficients of association can be used to assess *external validity*. [emphasis added] (p. 5)

The APA Task Force also stated (1) "In the absence of randomization, we should do our best to investigate sensitivity to various untestable assumptions" (p. 4); (2) "Describe any anticipated sources of attrition due to noncompliance, dropout, death, or other factors" (p.

6); (3) "Describe the specific methods used to deal with experimenter bias, especially if you collected the data yourself" (p. 4); (4) "When you interpret effects, think of credibility, generalizability, and robustness" (p. 16) ; (5) "Are the design and analytic methods robust enough to support strong conclusions?" (p. 16); and (6) "Remember, however, that acknowledging limitations is for the purpose of qualifying results and avoiding pitfalls in future research" (p. 16). It could be argued that these six statements pertain to validity. However, the fact that internal and external validity was not directly mentioned by the ASA Committee on Professional Ethics, as well as the fact that these concepts were mentioned only once by the APA Task Force and were not directly referenced in the "Discussion" section of the its report, is a cause for concern, bearing in mind that the issue of internal and external validity not only is regarded by instructors of research methodology, statistics, and measurement as being the most important in their fields, but that it also receives the most extensive coverage in their classes (Mundfrom, Shaw, Thomas, Young, & Moore, 1998).

In experimental research, the researcher manipulates at least one independent variable (i.e., the hypothesized cause), attempts to control potentially extraneous (i.e., confounding) variables, and then measures the effect(s) on one or more dependent variables. According to quantitative research methodologists, experimental research is the only type of research in which hypotheses concerning cause-and-effect relationships can be validly tested. As such, proponents of experimental research believe that this design represents the apex of research. An experiment is deemed to be valid, inasmuch as valid

cause-effect relationships are established, if results obtained are due *only* to the manipulated independent variable (i.e., possess internal validity) and are generalizable to groups, environments, and contexts outside of the experimental settings (i.e., possess external validity). Consequently, according to this conceptualization, all experimental studies should be assessed for internal and external validity.

Undoubtedly the seminal works of Donald Campbell and Julian Stanley (Campbell, 1957; Campbell & Stanley, 1966) provides the most authoritative source regarding threats to internal and external validity. Campbell and Stanley identified the following eight threats to internal validity: history, maturation, testing, instrumentation, statistical regression, differential selection of participants, mortality, and interaction effects (e.g., selection-maturation interaction) (Gay & Airasian, 1999). Additionally, building on the work of Campbell and Stanley, Smith and Glass (1987) classified threats to external validity into the following three areas: population validity (i.e., selection-treatment interaction), ecological validity (i.e., experimenter effects, multiple-treatment interference, reactive arrangements, time and treatment interaction, history and treatment interaction), and external validity of operations (i.e., specificity of variables, pretest sensitization).

Although experimental research designs are utilized frequently in the physical sciences, this type of design is not as commonly used in social science research in general and educational research in particular due to the focus on the social world as opposed to the physical world. Nevertheless, since Campbell and Stanley's conceptualization, many researchers have argued that threats to internal and external validity not only should be

evaluated for experimental designs, but are also pertinent for other types of quantitative research (e.g., descriptive, correlational, causal-comparative, quasi-experimental). Unfortunately, with respect to non-experimental quantitative research designs, it appears that the above sources of internal and external validity do not represent the realm of pertinent threats to the validity of studies. Moreover, Onwuegbuzie (2000a) contends that threats to internal and external validity should be assessed comprehensively in *all* quantitative research studies, regardless of the research design. Onwuegbuzie (2000a) provided a more comprehensive framework of dimensions and sub-dimensions of internal and external validity. Newly-conceptualized threats to validity identified by Onwuegbuzie included *observational validity, behavior bias, participant augmentation, treatment duration, restriction in range of measurement*, and *analytical errors* (e.g., model mis-specification, Types I-IV errors, non-consideration of effect size).

As noted by Onwuegbuzie (2000a), a paucity of researchers provide a commentary of threats to internal and external validity in the discussion section of their articles. Thus, journal reviewers and editors should strongly encourage all manuscripts to include a discussion of the major rival hypotheses in their investigations. In order to motivate researchers to do this, it must be made clear to them that such practice would improve the quality of their paper, not diminish it. Indeed, future revisions of the *American Psychological Association Publication Manual* (APA, 1994) should provide *strong* encouragement for all research reports to include a discussion of threats to internal and external validity. Additionally, the Manual should urge researchers to furnish a summary of the major threats

to internal and external validity for some or even all of the studies that are included in their reviews of the related literature.

Once discussion of rival hypotheses becomes commonplace in literature reviews, *validity meta analyses* could be conducted to determine the most prevalent threats to internal and external validity for a given research hypothesis (Onwuegbuzie, 2000a). These *validity meta analyses* would provide an effective supplement to traditional meta analyses. In fact, the *validity meta analyses* could lead to *thematic* effect sizes being computed for the percentage of occasions in which a particular threat to internal or external validity is identified in replication studies (Onwuegbuzie, 2000b). For example, a narrative that combines traditional meta analyses and *validity meta analyses* could take the following form:

> Across studies, students who received Treatment A performed on standardized achievement tests, on average, nearly two-thirds of a standard deviation (Cohen's (1988) *Mean d* = .65) higher than did those who received Treatment B. This represents a large effect. However, these findings are tempered by the fact that in these investigations, several rival hypotheses were noted. Specifically, across these studies, *statistical regression* was the most frequently identified threat to internal validity (prevalence rate/effect size = 33%), followed by *mortality* (effect size = 22%). With respect to external validity, *population validity* was the most frequently cited threat (effect size = 42%), followed by reactive arrangements (effect size = 15%)....

Such *validity meta analyses* would help to bolster further the importance of external

replications, which are the essence of science (Onwuegbuzie & Daniel, 1999; Thompson, 1994).

Currently, there are several websites that attempt to help students to become familiar with and to apply Campbell and Stanley's threats to internal and external validity. These websites tend to provide vignettes and then ask the student to choose from a list of 3-5 options the threat to validity that is most salient. Although these websites are useful, they raise five major concerns. First, as stated above, the websites subsume discussion of these threats under experimental designs, thereby giving the impression to some students that threats to internal and external validity are not an issue for other types of quantitative research. Thus, a recommendation is to utilize a broader conceptualization of the threats to internal and external validity in designing web-based tools for learning about validity, such as that proposed by Onwuegbuzie (2000a).

Second, the fact that these vignettes represent multiple-choice formats also gives the impression that each research study has only one threat to internal or external validity, an inaccurate assumption. Therefore, it is suggested that vignettes be created that have several rival hypotheses for each scenario provided. Third, in receiving immediate feedback (i.e., solutions), some students may not reflect deeply enough about the scenarios, preferring to select a response hastily in order to obtain early validation. In such cases, the critical thinking process involved in the rival hypothesis reasoning will be stunted. Interestingly, critical thinking has been found to be positively related to performance in research methods classes (Onwuegbuzie, Schwartz, & Rice, 2000).

Moreover, providing immediate feedback for web-based users is not compatible with the constructivist view of learning. Consequently, it is recommended that students are encouraged to develop vignettes either informally or formally (i.e., as part of a course assignment). These vignettes could then be posted locally or even globally. With respect to the latter, links could be made to other research tools such as statistical tutorials.

Moreover, special website research validity "chatrooms" could be set up throughout the United States and, subsequently, the world to discuss the rival hypotheses pertinent to a variety of scenarios. Indeed, these chatrooms could be utilized to facilitate the development of new internal and external validity categories. Such development is justified because research is not a static field, but one which continually evolves.

Fourth, although analyzing vignettes is an extremely useful exercise, it should be remembered that these vignettes represent mere isolated fragments of information, typically devoid of any theoretical framework. Thus, graduate students also should be taught how to critique full-length published research articles using an internal/external validity categorization scheme. This can be facilitated by posting some of these articles on the web (after obtaining author/editor permission) and then asking students to identify the possible rival hypotheses. By reading the entire article students will then be able to put assessments of validity threats in their proper context. Also, by setting up an open-response format for identifying rival hypotheses in these posted studies, instructors of research methodology courses could assess students' responses to determine the types of misconceptions they have about internal and external validity.

Finally, providing only web-based tools for learning about the validity of empirical studies will give graduate students and researchers alike the false impression that validity is not an issue in qualitative research. Indeed, although Huck and Sandler's (1979) book, *Rival Hypotheses: Alternative Explanations for Data-Based Conclusions*, focuses on empirical research, there is no reason why qualitative research cannot be included--bearing in mind that (1) qualitative research also generates *data*; and (2) most introductory educational research courses provide students with exposure to both quantitative and qualitative techniques.

As noted by Onwuegbuzie and Daniel (2000), a serious and consistent analytical error made by qualitative researchers includes a failure, often for philosophical reasons, to legitimize research findings and interpretations by providing an assessment of validity (e.g., credibility, relativism, external criticism). Unfortunately, although the importance of validity has long been accepted in the quantitative research community, the issue of validity has been controversial among qualitative researchers. In fact, qualitative researchers are divided as to whether validity should play a role in their discipline.

At one end of the qualitative continuum are those (e.g., Miles & Huberman, 1984) who believe that validity for qualitative research should be defined in much the same way as it is for quantitative research. According to this school of thought, internal validity and external validity should be assessed in qualitative research studies in a manner similar to that in empirical studies. At the other end of the spectrum, many post-modernists (e.g., Wolcott, 1990) contend that validity cannot and should not be assessed in qualitative

research because the researcher serves as the instrument. These individuals maintain that social processes are unpredictable, interactive phenomena which cannot be separated from the researchers' ways of identifying and interpreting them. Accordingly, researchers' observations are mind-dependent, with each interpretation representing just one of multiple realities in existence; consequently, the validity of qualitative research cannot be assessed. Moreover, many relativists define validity as representing whatever the community agrees it should represent. Such a definition is extremely vague, as well as being counterproductive because it misleads graduate students into adopting an "anything goes" mindset about qualitative research (Onwuegbuzie & Daniel, 2000).

As asserted by Onwuegbuzie (2000c), in order to be taken seriously, qualitative researchers must be held accountable for their data collection, analysis, and interpretive approaches. This can only be accomplished by providing evidence of representation and legitimization. According to Onwuegbuzie (2000d), many qualitative researchers reject the concept of validity because of their perceptions that the positivist framework (i.e., correspondence of truth) of validity often is utilized as the standard against which all other standards are conceptualized and assessed. As a result, they believe that in order to reject positivism, they must reject validity (Onwuegbuzie, 2000c). However, this should not be the case. Instead, concepts associated with the quantitative research paradigm such as internal and external validity should be avoided, so as to prevent such a reactionary view of validity among qualitative researchers, and an alternative framework for validity in qualitative research should be adopted.

As posed by Onwuegbuzie and Daniel (2000), if qualitative research cannot be assessed for validity and there can be no standards for this line of inquiry, then how is it that editors of qualitative journals such as *Qualitative Studies in Education* can determine which studies are published? Moreover, if qualitative research studies cannot be evaluated with respect to validity, then why do we need to teach qualitative research methodologies in graduate programs, since, presumably, any qualitative research that they conduct will be valid? Extending this argument further, surely a reader would have more confidence in the findings of a qualitative research study if the data emerged from a classroom observation that lasted the whole lesson (i.e., *prolonged engagement*) rather than from one that lasted for only the first few minutes of the class session? Similarly, surely a reader would find data more trustworthy if they emerged from several classroom observations (i.e., *persistent observation*) rather than from one?

Thus, in order for qualitative research to maximize its credibility in the educational research field, more rigor is needed (Onwuegbuzie, 2000c; Onwuegbuzie & Daniel, 2000). To this end, it is imperative that qualitative researchers assess the truth value of their findings. This can be accomplished by re-framing the concept of validity in qualitative research, for example, by treating validity as an issue of choosing among rival interpretations and of examining and providing arguments for the relative credibility of competing knowledge claims (Polkinghorne, 1983), or by re-defining validity as having multi-faceted criteria (e.g., credibility, transferability, dependability, confirmability; Lincoln & Guba, 1985).

For example, Onwuegbuzie (2000c) used Creswell's (1998) five-design conceptualization of qualitative research (i.e., historical, case study, ethnographic, phenomenological, and grounded theory) to develop a comprehensive list and description of methods for assessing the truth value of qualitative research. Such techniques included triangulation, prolonged engagement, persistent observation, leaving an audit trail, member checking, weighting the evidence, checking for representativeness of sources of data, checking for researcher effects, making contrasts/comparisons, checking the meaning of outliers, using extreme cases, ruling out spurious relations, replicating a finding, assessing rival explanations, looking for negative evidence, obtaining feedback from informants, peer debriefing, clarifying researcher bias, and thick description. Utilizing and documenting such techniques should help to reduce methodological errors in qualitative research (Onwuegbuzie & Daniel, 2000). Additionally, as noted by Constas (1992, p. 255), unless methods for examining rival hypotheses in qualitative research are developed, "the research community will be entitled to question the analytical rigor of qualitative research"--where rigor is defined as the attempt to make data and categorical schemes as public and as replicable as possible (Denzin, 1978). Thus, any future development of validity categories must consider both quantitative and qualitative research designs.

As a final note, generating and sustaining the use of rival hypotheses web-based tools will be time-consuming for research methodology teachers. Thus, in order to motivate instructors to adopt such tools, some recognition is needed of this teaching aid on the part of administrators of tertiary institutions. Moreover, it is likely that changes in the reward

structure will facilitate the growth of this and other web-based tools. For example, incorporating an effective validity web-based tool into lesson plans could be given the same status (i.e., credit towards tenure and promotion) as does the publication of one or more published refereed articles, since the fruits of faculty labor could be communicated worldwide via a website. This, in turn, not only would provide the underlying institution with more exposure, but also would facilitate dialogue among different universities. Simply put, carefully aligning the reward structure to the development and sustenance of a web-based tool should enhance the appeal of teaching the concept of validity.

References

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.

Campbell, D.T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54*, 297-312.

Campbell, D.T., & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: John Wiley.

Constas, M.A. (1992).. Qualitative analysis as a public event: The documentation of category development procedures. *American Educational Research Journal, 29*, 253-266.

Creswell, J.W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage Publications.

Denzin, N. (1978). *Qualitative analysis for social scientists*. New York: Cambridge University Press.

Gay, L.R., & Airasian, P.W. (1999). *Educational research: Competencies for analysis and application* (6th ed.). Englewood Cliffs, N.J.: Prentice Hall.

Huck, S.W., & Sandler, H.M. (1979). *Rival hypotheses: Alternative interpretations of data based conclusions*. New York: Harper Collins.

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Miles, M.B., & Huberman, A.M. (1984). Drawing valid meaning from qualitative data: Toward a shared craft. Educational Researcher, 13, 20-30.

Mundfrom, D.J., Shaw, D.G., Thomas, A., Young, S., & Moore, A.D. (1998, April). *Introductory graduate research courses: An examination of the knowledge base*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Onwuegbuzie, A.J. (2000a). *Expanding the framework of internal and external validity in quantitative research*. Unpublished manuscript.

Onwuegbuzie, A.J. (2000b). *Effect sizes in qualitative research: A prolegomenon*. Unpublished manuscript.

Onwuegbuzie, A.J. (2000c). *Validity and qualitative research: An oxymoron?* Unpublished manuscripts, Valdosta State University, Valdosta, Georgia.

Onwuegbuzie, A.J. (2000d). *Positivists, post-positivists, Post-structuralists, and post-modernists: Why can't we all get along? Towards a framework for unifying research paradigms*. Unpublished manuscripts, Valdosta State University, Valdosta, Georgia.

Onwuegbuzie, A.J., & Daniel, L.G. (1999, November). *Uses and misuses of the correlation coefficient*. Paper presented at the annual conference of the Mid-South Educational Research Association, Point Clear, AL.

Onwuegbuzie, A.J., & Daniel, L.G. (2000, April). *Common analytical and interpretational errors in educational research*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Onwuegbuzie, A.J., Schwartz, R.A., & Rice, D.(2000). *The role of critical thinking in educational research courses*. Manuscript in preparation.

Polkinghorne, D. (1983). *Methods for the human sciences*. Albany, New York: University of New York Press.

Smith, M.L., & Glass, G.V. (1987). *Research and evaluation in education and the social sciences*. Englewood Cliffs, NJ: Prentice Hall.

The American Statistical Association. (1999). Ethical guidelines for statistical practice [On-line]. Available: *http://www.amstat.org/profession/ethicalstatistics.html*

Thompson, B. (1994). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. *Journal of Personality, 62*, 157-176.

Wilkinson, L. and the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.

Wolcott, H.F. (1990). On seeking--and rejecting--validity in qualitative research. In E.W. Eisner & A. Peshkin (Eds.), *Qualitative inquiry in education: The continuing debate* (pp. 121-152). New York: Columbia University, Teachers College Press.